

Object Detection In Surveillance Videos

Anshuman Biswal



Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela – 769 008, India

Object Detection In Surveillance Videos

Dissertation submitted in

May 2014

to the department of

Computer Science and Engineering

of

National Institute of Technology Rourkela

in partial fulfillment of the requirements

for the degree of

Master of Technology

by

Anshuman Biswal

(Roll 212CS1105)

under the supervision of

Prof. Banshidhar Majhi



Department of Computer Science and Engineering

National Institute of Technology Rourkela

Rourkela – 769 008, India

dedicated to my Parents...



Computer Science and Engineering
National Institute of Technology Rourkela

Rourkela-769 008, India. www.nitrkl.ac.in

Dr. Banshidhar Majhi

Professor

May , 2014

Certificate

This is to certify that the work in the thesis entitled *Object Detection in Surveillance Videos* by *Anshuman Biswal*, bearing roll number 212CS1105, is a record of an original research work carried out by him under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of *Master of Technology* in *Computer Science and Engineering*. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Dr. Banshidhar Majhi

Acknowledgement

This dissertation, though an individual work, has benefited in various ways from several people. Whilst it would be simple to name them all, it would not be easy to thank them enough.

The enthusiastic guidance and support of *Prof. Banshidhar Majhi* inspired me to stretch beyond my limits. His profound insight has guided my thinking to improve the final product. My solemnest gratefulness to him.

I am also grateful to *Asst. Prof. Pankaj Kumar Sa* for his ceaseless support throughout my research work.

It is indeed a privilege to be associated with people like *Prof. S. K. Rath, Prof. S.K.Jena, Prof. D. P. Mohapatra, Prof. A. K. Turuk, Prof. S.Chinara* and *Prof. B. D. Sahoo*. They have made available their support in a number of ways.

Many thanks to my comrades and fellow research colleagues. It gives me a sense of happiness to be with you all. Finally, my heartfelt thanks for the unconditional love and support of my parents. Words fail me to express my gratitude towards them who sacrificed their comfort for my betterment.

Anshuman Biswal

Abstract

In this thesis, a novel scheme for object detection in complex background scenes has been proposed. The input videos used have fixed backgrounds and static cameras. Initially median of few frames is evaluated for obtaining a proper estimate of the background. Local threshold based background subtraction is done for extracting objects from the video sequence. During sudden illumination changes, optical flow analysis is used for motion segmentation. It is assumed that during photometric distortions, the object is in motion. Subsequently shadow detection and suppression is done to the resulting thresholded image. Hue Saturation Value(HSV) color space model is used for shadow suppression. Visual measures convey the performance of the algorithm.

Keywords: HSV color space, optical flow, object detection, shadow suppression.

Contents

Certificate	iii
Acknowledgment	iv
Abstract	v
List of Figures	viii
1 Introduction	2
1.0.1 Object Detection	4
1.0.2 Optical Flow	8
1.0.3 Color Space Models	9
1.0.4 Generic Tracking Model	11
1.1 Research Goal	12
1.2 Thesis Layout	13
2 Literature Survey	15
2.1 Related work	16
3 Proposed Work	31
3.1 Background	31
3.2 Proposed Detection Model	32
3.2.1 Background Modelling	32
3.2.2 Background Subtraction	32
3.2.3 Motion Segmentation using Optical Flow	34

3.2.4	Shadow Removal using HSV color space	34
3.3	Results	37
4	Conclusions and Future Scope	45
4.1	Scope for Further Research	46
	Bibliography	47

List of Figures

1.1	Memory view of the video frames in a computer	3
1.2	RGB color space model	10
1.3	HSV color space model	10
1.4	Generic Detection Model	11
3.1	Proposed Detection Model	33
3.2	Shadow Suppression of <i>VipTraffic</i> image. (a) True Image, (b) Thresholded Image (c) Shadow suppressed	39
3.3	Shadow Suppression of <i>VipTraffic</i> image. (a) True Image, (b) Thresholded Image (c) Shadow suppressed	40
3.4	Shadow Suppression of <i>VipTraffic</i> image. (a) True Image, (b) Thresholded Image (c) Shadow suppressed	41
3.5	Input frames of <i>Traffic</i> sequence containing sudden illumination change	42
3.6	Inability of the existing model to extract objects accurately	42
3.7	Object extraction using optical flow thresholding	43
3.8	Original rgb frames of the Light Switch sequence	43
3.9	Inability of the existing model to detect object	44
3.10	Result of optical flow analysis	44

Chapter 1

Introduction

In the current generation of rapidly developing technologies, multimedia has deeply penetrated into all realms of life. One's daily routine has multiple encounters with multimedia services. One prominent reason for sudden upsurge of multimedia components is decrease in cost of technological gadgets like cameras and computers. Cameras have evolved rapidly throughout the century. The cost of image sensors has reduced drastically resulting in abundance of imaging devices. This resulted in huge accumulation of data in form of images and video clips. Extracting relevant information from this multimedia content was highly important. This need led to the development of object detection algorithms. Object detection methods play a vital role in the domain of surveillance. Accurate object detection and tracking led to higher cognitive tasks like event classification. Object tracking comprises of two nearly related procedures; object detection followed by process in which the detected objects are tracked. Object tracking is basically estimating the object's position using past information about its motion. Object tracking algorithms have gained quite popularity due to inexpensive yet high quality cameras and growing demand for automatic video analysis. The surveillance system is the systematic method of behavior monitoring, actions or other changing information. Ideally the object detection system should have a static and uniform background from which the object may be extracted by using simple background subtraction operation. Object detection, object tracking and

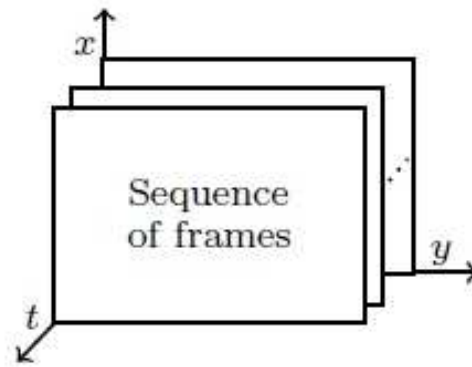


Figure 1.1: Memory view of the video frames in a computer

recognition are three fundamental steps in any surveillance system. Video analysis comprises of relevant moving object detection, tracking of corresponding object from frame to frame and study of the motion history image to study its behavior.

Videos comprise of image sequences called frames. Image conveys only spatial information, whereas video has spatial as well as temporal information of the scene. The processing of video clip for information extraction is called as digital video processing. In computer, images are stored as matrices whose values depict the intensity values of the pixel positions in the image. Each element in this matrix is referred to as pixel (picture element). An image could be defined as a two-dimensional function, $I(x, y)$, where x and y are the position co-ordinates, and the amplitude of I at a particular position is called the gray level or intensity of the image at that specific point [1]. Video may be defined as function of images over time, $R(x, y, t)$ where t represents the frame number. In other words, videos are represented as three dimensional matrices having two position components and one component for frame number. $R(x, y, t)$ denotes the gray scale value of the picture element in position (x, y) in the frame number t .

Here video segmentation technique is used for object extraction. Video segmentation is a systematic procedure of dividing the frames into temporally coherent regions. Object detection is one of the several practical applications of video segmentation. Here some key concepts used in the domain of video

segmentation applications are discussed.

1.0.1 Object Detection

Locating the first appearance of an object in a video clip is commonly called as object detection [2]. It can be done using a single frame, but a more robust method is to obtain information from multiple frames. This method reduces the number of false detections. Frame differencing is a popular way to obtain the temporal information from the video sequence. Changing regions in consecutive frames are highlighted using this method. Once these regions are obtained then they form the input for the object tracking module to keep record of the object's position efficiently across frames.

There are diverse approaches for object detection algorithms. Broadly they can be categorized into the following groups :

Point Detectors

Important points in images are those points which can express texture in their neighboring regions in a. The method used to find such points is called point detectors. These points should be vary with the scale and resilient to changes in illumination. They should not depend on the camera viewpoint. Most popular detectors are Moravecs interest operator [3], Harris interest point detector [4], KLT detector [5], and SIFT detector [6].

In Moravecs operator variation of image intensities across a 4x4 pixel grid is computed to find the interest points. The computation is done in horizontal, vertical, diagonal and anti-diagonal directions; then the least of the four variations is selected as the symbolic value for that window. The criteria for declaring a point as interesting is that its intensity variation should be locally maximum in a 12x12 patch.

Harris interest point detector uses the first order image derivatives along x and y directions to evaluate the variation of values along specific directions. Consequently this variation is encoded by a secondary moment matrix which is

computed for every pixel in the small locality. The determinant and trace of this matrix is used to decide whether a point is interesting or not.

KLT(KanadeLucasTomasi) method also uses the same moment matrix for defining the interestingness of a point. Eigenvalues of the matrix has to be computed to get the interest point confidence values. Thresholding of this confidence value is done to get the relevant interesting point.

The moment matrix is generally robust to rotary motion and change of location. Though it changes with respect to affine or projective transformations. Scale-invariant feature transform (SIFT) detector is resilient to such transformations. This generates much more number of relevant points as compared to other alternatives. Main reason behind this is that the interest points are gathered across different scales and diverse resolutions. SIFT is more robust to image distortions as compared to other detectors.

Background Subtraction

Another technique for object detection is background subtraction. The idea is to relate the static background frame against the current video frame containing the moving object pixel by pixel. This involves the initial development of the background model. Then subsequently the incoming frames are compared with the model to find the regions of significant changes. Generally, a connected component method is further applied to obtain regions which are not disjoint corresponding to the objects. This method is popularly known as the background subtraction [2].

Segmentation

The objective of image segmentation methods is to separate the image into perceptually related areas. These methods address two issues, the condition for a good segmentation and the procedure for obtaining good division. Some of the popular segmentation techniques are stated below .

Mean-Shift clustering was developed by Comaniciu and Meer. This algorithm

tries to find groups in the collective spatial and color space (l, u, v, x, y) , where (l, u, v) represents the chromaticism and (x, y) represents the position co-ordinates [7]. Initially after taking the input image, initialization is done with big number of virtual cluster centers arbitrarily chosen from the samples. After that each cluster centroid is shifted to mean of the samples which are positioned internally to the multidimensional ellipsoid centered on the centroid. Mean shift vector is the variable quantity characterized by the old and current centroids. This process is repeated till the centroids don't change their positions. Meanwhile during the process merging of clusters may take place. This method is scalable to different applications like edge detection and image regularization. Adjustment of several variables has to be done for optimal performance. The variables include choice of the chromatic properties and spatial kernel bandwidths, and the limit for the least size of the area considerably effects the resulting partition.

Graph cuts : In a typical graph partitioning problem the vertices of the graph are divided into N disjoint regions(subgraphs), by limiting the prioritized links of the graph. Segmentation problem can be formulated in similar fashion. The vertices are replaced by pixels, and the graph is substituted by the entire frame. The goal is to divide the whole frame into disjoint regions. The complete priority of the limited links between two subgraphs is defined as a cut. Priority is generally evaluated by using any important feature like chromatic characteristics, intensity or texture relationship among the nodes. One drawback of this method is that they are quite costly in terms of memory and computation needs. The advantage of this scheme is that it requires less hand picked variables as compared to other technique stated above.

Active contours : In this method, partition is done by developing a limited contour to the element's extremity in a fashion that the spatial property tightly bounds the element's region. An energy functional which states the contour suitability to the virtual object area governs the evolution of the contour. A vital point in this method is the contour initialization. A contour is generally located externally the object area and diminished until the object boundary is reached in a image gradient based approach. In region based methods this constraint is

relaxed so that the contour can shrink or expand depending on the initialization of the contour is done inside or outside. The problem with this approach is that it requires prior object knowledge. This issue can be resolved by using several frames or a indicator frame in which initialization could be done without construction region priors.

Supervised classifiers

Supervised learning mechanism involves learning diverse object perceptions naturally from a training set. Complete set of templates has to be stored as a requirement for proper learning. Given a suitable training set, these schemes develop a method that links input to suitable output. Supervised learning consists of regression or classification. Regression is the method in which the element which learns estimates the function's activity by generating a continuous value. If instead of a continual value, label is generated then the method is called classification. The training set contains combination of object characteristics and category label in which both the amounts are mechanically defined. The main criteria which decides the performance of this method is the selection of good features. All features won't necessarily give the best performance. The selection criterion of features depends on the discriminating power of the feature which represents how strongly it differentiates one class from another. Upon selection of the features , a specific learning approach has to be considered for proper working of the system. There are several popular learning approaches like neural networks, adaptive boosting, support vector machines and other techniques. Two popular approaches used for object tracking are adaptive boosting and support vector machines.

Adaptive Boosting: It is an repetitive scheme of searching a very efficient allocator by collecting multiple fundamental allocator, which are partly efficient [8]. During the training phase, the first step is the construction of an introductory arrangement of weights over the samples. Subsequently boost method chooses the fundamental allocator which gives the least fault, where fault is proportionate to the priorities of the incorrectly classified data. Subsequently weights related

with the misclassified sample by the chosen fundamental allocator are increased. Therefore the method supports the choosing of alternate allocator which operates better on the incorrectly classified sample in subsequent repetition.

Support Vector Machines(SVM) : This method is utilized to classify the available samples into binary categories by searching the maximal marginal hyperplane which separates the categories. The linear space betwixt the hyperplane and the nearest sample is called as the margin of the hyperplane which is stretched. The sample points which are located on the margin limitation of the hyperplane are called as support vectors. In scenarios of object detection, these categories correspond to the positive samples(object classes) and negative samples(non object classes). Quadratic programming is used to compute the hyperplane from multiple possible hyperplanes from the manually generated training set. Though SVM is mainly a linear classifier, it can be adapted to forms a non-linear classifier using the kernel trick. This kernel projects the data into higher dimensional space where the data becomes linearly separable .But the selection of a suitable kernel is quite tricky. After choice of a kernel, exhaustive testing has to be done for testing the classification performance which might not work when fresh observations are added to the sample set.

Here in this thesis, background subtraction method is used because of the ease of computation and less memory requirements.

1.0.2 Optical Flow

This method helps in representing the motion between two frames by the use of motion vectors. They help in producing dense flow fields by computation of the flow vector of individual picture element subjected to the brightness constancy constraint [2]. Brightness constancy constraint states that the gray scale value of a picture element does not vary very much across two frames.

$$I(x, y, t) - I(x + dx, y + dy, t + dt) = 0 \quad (1.1)$$

Here in the above equation, $I(x, y, t)$ refers to the gray scale value of the pixel at location (x, y) in the frame number t . This computation of optical flow is done in the neighborhood of the pixel. There are two popular methods for optical flow calculation namely Horn and Schunk method [9] and Lucas Kanade method [10].

1.0.3 Color Space Models

The objective of a color model is to promote the characterization of colors in specific method [1]. Essentially it is a method of representing a single color by a single point in specific subspace within a particular coordinate system. They can also be seen as virtual mathematical paradigm in which chromatic characteristics are portrayed as numeric pairs. There are several popular existing color models. Here in our thesis we are using two popular models namely Red Blue Green (RGB) color space and Hue Saturation Value (HSV) color space.

RGB color model

In this model, each color component exists in its basic spectral segments of blue, green and red. Cartesian coordinate system is the base for this model.

The color subspace is portrayed by the cube as shown in the figure 1.2. Cyan, magenta and yellow are located at the three corners. Black is located in the origin. White is located diagonally opposite to black where the values of all color components are maximum. The various colors defined by this paradigm are points lying on the cube surface or internal to it. Gray scale which represents points having equal values of three components lies on the line joining black and white color point.

HSV color model

RGB color space model is not well suited for describing colors in terms for human interpretation. Humans don't perceive colors as composition of three primary color components. HSV is rather a more intuitive color space model which represents colors in the way human perceive it. H stands for Hue, S stands for saturation

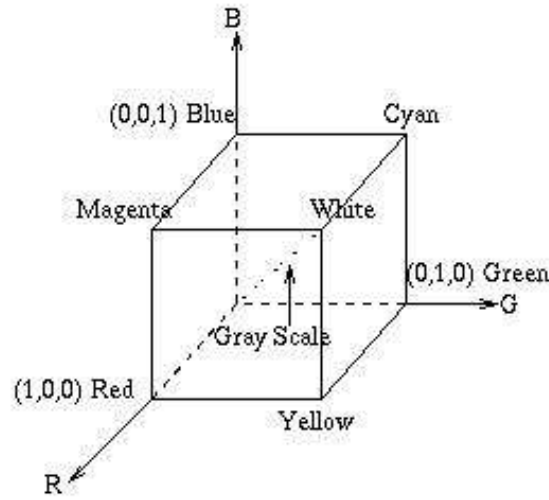


Figure 1.2: RGB color space model

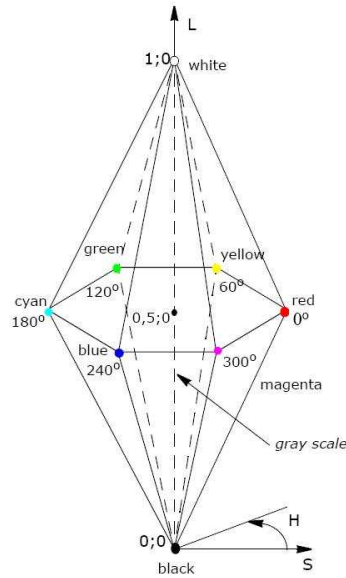


Figure 1.3: HSV color space model

and V stands for Value. Hue represents the chromatic parameter which represents the principal wavelength in a mixture of light waves [1]. Basically it exhibits the principal chromatic feature as sensed by the human eye. The purity of the color is represented by its saturation value. This reflects the quantity of white light combined with the hue. Fully saturated colors are the pure spectrum colors. Value

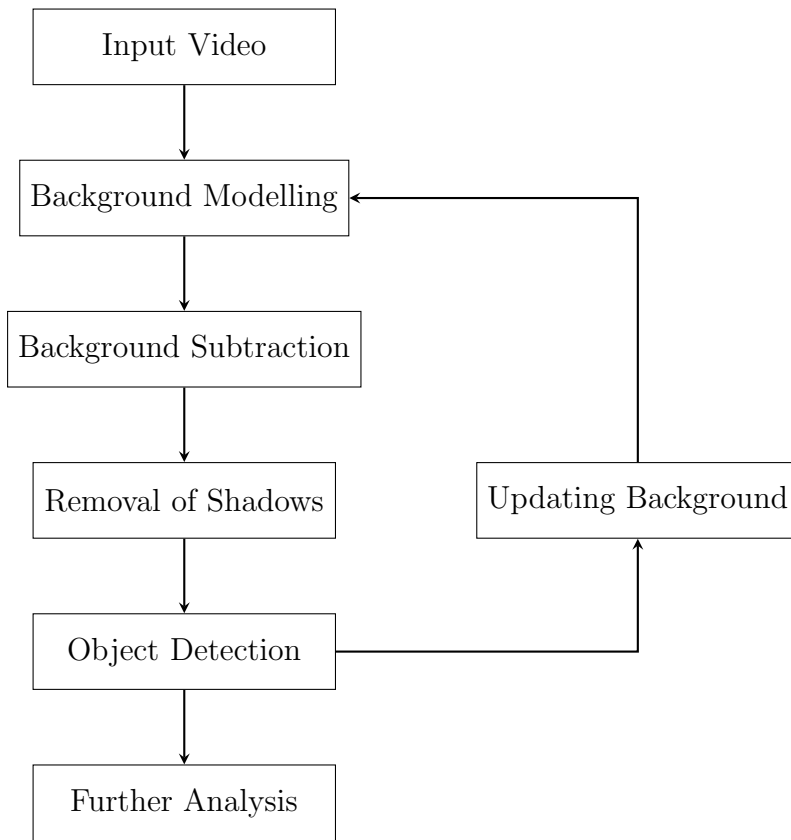


Figure 1.4: Generic Detection Model

represents the brightness of the color. It is similar to the amount of light coming out from the color.

1.0.4 Generic Tracking Model

During movie recording camera position and scene dynamics have to be seriously considered for proper recording. Camera position may be static or dynamic. There may be camera shake. Scene may be static or dynamic. Scene may be a simple one or a cluttered one. The lighting conditions may be uniform or changing gradually. There may be sudden illumination changes.

In this thesis input videos are obtained from fixed camera position and static background.

The above mentioned criterion makes background subtraction method suitable for object detection in our case. Initially a background model is developed using

initial few frames. The background model so modeled forms a suitable estimate of the actual background which exists in the scene. Subsequently incoming frames will be compared with the existing background model pixel by pixel basis to extract the moving foreground objects. Unfortunately shadows are also falsely detected as foreground objects. Separate shadow suppression methods have to be implemented to get the actual moving object regions. The model may be updated in order to accommodate changes in the background scene.

This generic model works in ideal conditions, but fails miserably in real life conditions. The major challenges faced by this model are as follows :

- Periodic motion of background objects like waving of leaves, flow of water, tides, fountain, clouds
- Occlusion
- Gradual change in illumination
- Sudden change in illumination
- Shadows falsely detected as objects
- Object which stays stationary for some time is treated as background
- If stationary objects are moved from one place to another, the place where the actual object was existing is falsely detected as objects
- Adaptation of the model to the changing conditions is time consuming
- Glare is sometimes treated falsely as objects

1.1 Research Goal

After a thorough study of the advantages and drawbacks of the generic tracking model, our research goals are directed as follows:

- Development of a robust model which can accurately detect object during periodic motion of background components like waving of leaves, flowing of water or water movement in a fountain.
- Using a thresholding technique which suppresses the shadows to extract the actual moving objects.
- Even after the thresholding technique , a separate shadow suppression module is introduced to make the model more robust to strong shadows.
- Feature used should be invariant to complicated object motion or shape.
- Making the model more robust to sudden illumination changes by exploiting the motion of the object using optical flow information

1.2 Thesis Layout

The rest of the thesis is organized as follows :

Chapter 2 : Literature Survey The literature surveys that have been done during the research work in analyzing the issue of background subtraction in challenging situations is discussed here. It also provides a detailed survey of the literature related to motion detection, shadow removal and background subtraction .Discussion about the existing and recent methods for object detection is done.

Chapter 3: Proposed Work Here the proposed work has been discussed. The pre-existing local threshold based background subtraction method is modified to become more resilient to shadows and sudden illumination changes. Motion segmentation using optical flow feature is done to extract the moving regions during sudden illumination change. Shadow removal module is introduced which uses the HSV color space to suppress the strong shadows so that they are not falsely classified as objects.

Chapter 4: Conclusion This chapter provides the concluding remarks with a stress on achievements and limitations of the proposed schemes. The scopes for

further research are outlined at the end.

Chapter 2

Literature Survey

Detecting instances of semantic objects of a certain class in digital images and videos is generally referred as object detection. It has numerous applications in many broad areas of computer vision which includes pose estimation, image retrieval and video surveillance etc. Object detection is done in videos in which the camera may be static or dynamic. If the background is fixed and the camera is static, then background subtraction technique is a suitable technique for object detection. In such approach, moving objects are extracted from the scene by comparing the current frame from an already built background model.

Generally in most of the schemes, the object detected is also grouped together with incorrectly classified objects due to illumination variation., periodic motion of background elements, occlusion and various other reasons. One major drawback of such schemes is that shadows are falsely detected as objects. The pre-existing local thresholding based background subtraction algorithm reduced shadows which were faint [11]. However, it failed to remove strong shadows which occurred in video sequence where there was broad daylight. To solve this problem, a separate shadow removal module was introduced into the algorithm which uses the HSV color space to suppress the shadows. LIBS was unable to detect object during sudden illumination change. Most of the pixels were falsely classified as foreground objects during such event. This problem was alleviated by extracting the moving regions from the scene using the optical flow feature. It is assumed that the

objects are constantly in motion during the sudden change of illumination so that they can be detected using motion segmentation. During the development of the background model, the nature of each pixel is decided as stationary or not. Subsequently only stationary pixels are considered for the background model formation. In the background model, for each pixel location a range of values are defined. Then in object extraction phase our scheme employs a local threshold, unlike the use of global threshold in conventional schemes.

There are several existing schemes for background subtraction. We have covered few popular techniques in our literature survey which are described as below.

2.1 Related work

In the initial days, filters were popularly used for background modeling and subtraction. Koller et al. addressed the problem of multiple car tracking with occlusion reasoning in the similar fashion [12]. A contour tracker based on intensity and motion of boundaries was employed. Linear Kalman filter was used to achieve this. Kalman filter was used in more than one way, one for motion parameters estimation and another for predicting the shape of car's contour. One of the main aspect of background subtraction and modeling is the continuous maintenance of the background model so that it can adapt to the gradually changing background components. A three component system was developed by Toyoma et al. for background maintenance [13]. The three component were pixel level component, region level component and frame level component. Wiener filtering was performed by the pixel level component to make probabilistic estimation of the expected background. Region component was given the role to fill in uniform areas of foreground objects. Abrupt and global changes were taken care of by frame level component. Problems which occur at various scales is taken care of by this algorithm because of the three component system. Preliminary classification of foreground versus background is done by the first component processing which has the additional ability to adapt

variable backgrounds. The pixel level processing which happens at each pixel happens independently irrespective of the information observed in the neighboring pixels. Inter-pixel relationships are considered in the region level processing. This further refines the classification done at the pixel level, thereby alleviating the foreground aperture problem. Frame level processing takes care of the global illumination changes which may be falsely classified as objects. When there is sudden illumination change in large part of the frame, alternate background models are swapped which could explain the new background as much as possible.

Modelling the background severally at each pixel position (i, j) was proposed by Wren et al. [14]. Gaussian probability density function computation over the last n pixel values forms the basis of this model. Running average method is used to avoid the probability density function calculation from beginning at each new frame. The running average at time t is computed as follows,

$$\mu_t = \alpha I_t + (1 - \alpha)\mu_{t-1} \quad (2.1)$$

where I_t is the pixels present value, μ_{t-1} is the previous average, and α is an empirical weight. Standard deviation σ_t , which is the other parameter of a typical Gaussian probability density function can also be computed in a similar fashion. This method has the advantage of having low time complexity as well as space complexity because for each pixel only two parameters (mean, standard deviation) is stored as compared to the need of storing entire n values in additional memory. Then at each frame, pixel may be classified as foreground if its difference from the computed mean exceeds by some multiple of the standard deviation.

Koller et al., identified that the updation of the mean has to be more frequent [15]. So in his paper, he modified the above stated equation to as follows:

$$\mu_t = M\mu_t + (1 - M)(\alpha I_t + (1 - \alpha)\mu_{t-1}) \quad (2.2)$$

where the variable M can take only two values. Zero value signifies the presence of background, else it relates to foreground.

A different approach using median was used for background updation by Lo and Velastin [16]. Median value of last n frames was used as background

model. Variance filter was used for extracting features of objects which was robust to illumination variation. Cucchiara et al. explained that such a median value gives a suitable background paradigm even though the n frames are sub sampled relative to the actual frame rate by a factor of 10 [17]. But the main drawback of a median-based approach is that, its implementation requires a memory with the latest pixels information. This is a memory intensive process.

Stauffer and Grimson modelled each pixel by a mixture of Gaussian [18]. This method is popularly known as Gaussian Mixture Model (GMM). Multiple Gaussian components was used to incorporate static differences in the background scene. Suppose the background has moving leaves as well as shadows during the background development phase. Then the intensities of shadow pixels would be represented by a Gaussian curve, while that of leaves would be represented by another Gaussian curve. When combined together, there will be multiple Gaussians with their corresponding weights. Here individual pixel is modeled differently by a combination of K Gaussian,

$$P(I_t) = \sum_{i=1}^K w_{i,t} \times \mathcal{N}(I_t; \mu_{i,t}, \Sigma_{i,t}) \quad (2.3)$$

Where K ranges from three to five.

Kernel Density Estimation technique was used to model the background distribution by Elgammal et al [19]. This method was used on the buffer having the last n background values to model its distribution. Kernel Density Estimation method guaranteed a smooth and continuous version of histogram [20]. The background probability density function is modeled as the accumulation of the Gaussian kernels which are centered in the most latest n background values,

$$P(x_t) = \frac{1}{n} \sum_{i=1}^n (x_t - x_i, \Sigma_t) \quad (2.4)$$

This Kernel Density Estimation method can easily handle situations where there is small motion in the background. In scenes where the background is cluttered and there is periodic motion like waving of leaves, flow of water etc; this algorithm is able to give good results. Basically this model predicts the likelihood of finding

pixel intensity value depending on the history of the grayscale values of the pixel. The advantage of this model is its quick adaptability to the changes in the scenes which enables it to be very vulnerable in moving target detection. Color information is also used by this model to suppress the detection of shadows as objects.

Seki et al. described a method supported on the belief that, neighbouring group of pixels of background should not have much variations in a period of time [21]. Though this assumption works for a uniform background, but it fails to work in scenarios where the background is cluttered. The main reason behind this is that this assumption does not hold true for the pixels which are located at the region boundaries.

Principal Component Analysis method has also been used for background modeling. Initially few samples are collected gradually and later are utilized in develop a principal component analysis (PCA) model. The criteria for deciding a block of new video frame to be background or foreground depends on the distance similarity between the image pattern and its reconstructions using PCA projection coefficients of eight-neighbouring blocks. Power and Schoonees used similar technique for background modelling [22]. The problem with their approach was that an updating mechanism was lacking to adapt the blocks over time. PCA reconstruction error was focused by Oliver et al. [23]. Similarly Independent Component Analysis was also used for background modelling by Tsai and Lai [24]. Independent component analysis was performed on serialized images taken from a training sequence. After the computation of the resulting vector, comparison of the new image was done with it for extraction of the foreground objects from a developed background image. The main advantage of this method was it was resilient to indoor illumination changes; but it was computationally intensive.

Lin et al introduced a two-level mechanism based classifier [25]. Initially the classifier determined whether the current input image block belonged to foreground or background. In second stage corresponding block wise updates were performed depending upon the outcome of the first stage. Global validations of the local updates to maintain the inter block consistency is used to improve the quality

of the solution.

Artificial neural networks were used by Maddalena and Petrosino to learn the motion patterns by the background model by self organization [26]. This method give good results in scenes containing complication scenarios like moving backgrounds, camouflage, gradual illumination variations. It does not have bootstrapping limitations. Shadows are also successfully suppressed by this method.

A simple and effective method for object detection was presented by Haritaoglu et al. called as the W4 model [27]. In this model; three values were used for each pixel, which represented the minimum and maximum intensity values. The third values described the maximum intensity difference between consecutive images in the training sequence. A statistical background model is used for detecting the foreground pixels. Pixel wise median filter is used initially for differentiating between the stationary pixels and the moving ones. Subsequently only stationary pixel are used for the development of the background model. This model however does not work in scenes containing sudden illumination changes and in scenes where shadows are present.

The background model proposed by Gutchess et al. had several conjecture of the background value at each pixel were generated by finding regions of non-changing intensity in the input video [28]. Subsequently optical flow data from the locality around the picture element is used to evaluate the likelihood of each hypothesis. The background representative is the most likely hypothesis. The W4 model was further improved to include shadow detection and removal by Jacques et al [29]. Normalized cross correlation technique was used for shadow removal. It is assumed that the ascertained value of the shadowed pixels is directly scalable to the incident light. In other words, shadow pixels are the scaled version of the related pixels in the background model. So they should be highly correlated with the background pixels as compared to the object pixels. This criteria is used for shadow detection. The ratio of intensities of the shadow pixels and the corresponding background pixels should also be nearly constant. This information is used in shadow elimination.

A novel background subtraction scheme with shadow identification was proposed by C.R. Jung [30]. Robust estimators were used in the training phase for modelling the background. Detection of foreground pixels in the valuation stage by done a fast test. Shadow identification and removal is done by a statistical paradigm fused with estimated geometrical characteristics. Isolated foreground pixels are filtered by using morphological operations.

A global background subtraction method called ViBe for video sequences was proposed by Barnich and Droogenbroeck [31]. In this model, each background pixel usually takes values from its preceding frames in same location or its neighbor. Then pixel based comparison is done which involves current pixel value to check belongingness of the pixel with the background. It modifies by selecting arbitrarily which parameters to replace from the background paradigm. This method has a different updating mechanism. The basic idea it to accumulate the past samples and to update the sample values by not considering the information about when they were added to the models. Background model initialization is done from the first frame itself. For this, it is assumed that the neighboring pixels have similar temporal distribution.

Fuzzy color histogram was used by Kim and Kim to develop proper background subtraction method for temporally changing texture scenarios [32]. This clustering based technique has a quality of vastly suppressing color deviations developed by background movement while still emphasizing mobile entities. The idea behind this paradigm is that chromatic deviations produced by background movements are vastly suppressed in a fuzzy manner. So this method is nonparametric. Local features are derived from fuzzy color histogram. Then the construction of the background model is reliably done by using the computation of similarity between local FCH features with an on-line update procedure.

An overlapping block-by-block approach for detection of foreground objects was used by Reddy et al. [33]. In this method, texture information is passed from each block through three classifiers which classify them as background or foreground. A probabilistic voting scheme is used at the pixel level for result integration. This results in the final segmentation. The presence of three classifiers

makes this scheme quite effective.

A local thresholding based background subtraction method was introduced by Hati [11]. Instead of using a global threshold, here two thresholds were defined for each pixel. If the pixel values of the current frame lied outside the range of the thresholds, the pixel was classified as foreground. The model used initial few frames for background modeling. In this stage, the pixel values across a window of frames were collected. If the variation of the intensities was within a threshold, then the pixel was classified as stationary. These classified stationary pixels were further used for the background model development. In this phase, minimum and maximum thresholds were calculated for each stationary pixel. This method was robust to faint shadows and small local motions in the background. The main drawback of this paper was that it could not handle strong shadows and sudden illumination change scenarios.

Pilet et al. tried to solve the challenge of sudden illumination change by using a statistical model which relied on illumination effects rather than pixel intensities [34]. The main rationale behind this model is that illumination effects generally alter complete regions as compared to individual pixels. Here the statistical background model is replaced by statistical illumination model. Significantly the proportionality of intensities of the stored background image and the current image in the triple channels is modeled as Gaussian Mixture Model. This relates the information that different regions of the scenes may be affected in diverse ways. This GMM is incorporated in a probabilistic framework. By this way, texture, color and background illumination cues are taken into consideration. The fundamental logic behind this method is that under the assumption of a static background, variations in intensities of the non-occluded pixels happen due to global illumination effects. Changes are located through out the frame. Typically complete regions of the image are affected as compared to the individual pixels. Therefore, they can be modeled using GMMs having fewer components. The number of components used in this method is two. Histograms of correlation and amount of texture are trained beforehand, so that the information related to illumination, color and texture cues can be utilized. Pixel color of occluded objects

is also modeled as a mixture of Gaussian and uniform distribution. Correlation between image patches in the model and the input image is utilized to decide whether the pixels are occluded or not. This is done because illumination changes preserve the texture information while occlusion radically changes it. Pixel independence is assumed for lessening the computational intensity. Each pixel has five features, namely red, green, blue values and normalized cross correlation and texturedness values. This method is robust to illumination changes as well as shadows.

An alternative algorithm tried to tackle the challenge of sudden illumination change by developing an illumination change model, a chromaticity difference model, and a brightness ratio model [35]. This scheme was proposed by Choi et al. A chromaticity difference model and a brightness ratio model is developed which estimates the intensity difference and intensity ratio of false foreground pixels. Illumination change model forms the base for these models. They also don't need off line training about illumination change. The illumination change model is based on Phong shading model. Probability distribution of false foreground pixels is estimated by the chromaticity difference model. Subsequently separation of the foreground pixels obtained by GMM into moving object pixels and candidate false foreground pixels. However these candidate false foreground pixels have the possibility of including moving object pixels which may have zero chromaticity difference. Further filtering by the brightness ratio model helps in eliminating these irrelevant pixels. Finally the actual moving object pixels are extracted by cascading the two processes for removing the false foreground pixels arising because of sudden illumination change.

Another novel scheme based on codebook model was proposed by Kim et al. [36] In this method, the background sample values are quantized into codebooks. These codebooks represent a compressed form of background model for a long image sequence. Structural background variation due to periodic-like motion is captured by this codebook over a long period of time under limited memory. A codebook is built for each pixel using one or more codewords. Clustering of the samples at each pixel is done based on a color distortion metric together with

brightness bounds. Samples are clustered into set of codewords. Hence formed clustered don't represent a single Gaussian or other parametric distribution. There could be multiple codewords for a pixel, even if that pixel has a normal distribution. Background encoding is done on a pixel by pixel basis. Current image is compared with the background model in terms of color and bright differences for detection. An incoming pixel is classified as background based on two conditions. Firstly the color distortion to some codeword is less than the detection threshold. Secondly its brightness does not exceed the brightness range of that codeword. Even if it does satisfy only a single criterion, it is classified as foreground pixel. The main logic behind this that background pixel values lie in the same line as principal axis of the codeword along with the minimum and maximum bound of brightness, since the difference is mainly due to brightness.

Vosters et al. combined eigenbackground and statistical illumination model to tackle the problem of sudden illumination change [37]. Eigenbackground is used for reconstructing the background frame. Improvement of foreground segmentation is done by statistical illumination model. Detection of reliable background pixels is done by on line spatial likelihood model. In the training phase, the eigenbackground model is trained from a training sequence which contains the challenging scenarios of sudden illumination change of the background with a non-moving camera. Reconstruction for each input image is done to make the corresponding background image by this operation mode. The following method has the advantage of being responsive to changes in background which includes sudden local and global light changes and dynamic backgrounds. One drawback of this method is that it cannot learn the shadows created foreground objects, because in the training frames objects are not present. This problem is solved by using a statistical illumination method which models the remaining background variations caused by shadows and highlights of foreground objects. Here a on line spatial likelihood model is used as compared to pre-learned spatial likelihood model. This model is updated by using reliable background pixel detection. Here it is assumed that the training sequence of images don't have foreground objects in it, so that later they don't cause errors in the

background reconstruction stage. Eigenspace model is computed individually for each channel individually rather than a common model for all channels. An online spatial likelihood method is preferred because the pre-learned method is unable to adapt to new scenarios and may fail completely during drastic illumination changes. Initially for each input image, detection of reliable background pixels is done. These pixels help in the further updation of the background distribution. For reliable background pixels detection it is assumed that the ordering of the pixels in textured areas will not be disturbed during heavy photometric distortions. Comparison of sign of pixel difference in a locality is done as a measure for order preservation. Normalized cross correlation is used as a measure for texture. Integral image tables are used for efficient calculation of normalized cross correlation values. This method has very good performance in most of the challenging scenarios. The main drawback of it is that moved background objects are treated as foreground objects forever. This occurs because the eigenspace background model does not incorporate the object's new location. This results in reconstruction errors. Another problem occurs when there are large foreground objects in the input frames. Inaccurate segmentation occurs because of the degradation errors which are spread over the entire reconstructed background.

Another scheme for background modelling specifically for complex scenes was proposed by Calderara et al. [38] Background bootstrapping, shadow suppression, ghost removal and selective update of the background model were handled by proper techniques in this algorithm. An estimate of background is developed from initial few frames of the input video sequence. This procedure is called as background suppression. A critical task for a good background model is its development in the bootstrapping process. This stage is the initial stage when the background suppression is performed. Generally conventional background modeling schemes assume that during the initial training frames of the input video sequence there is no foreground object. But this assumption may not hold true in real life scenarios like traffic surveillance. This method takes care of those scenarios. It is able to extract the background estimate from the initial frames which may contain the moving object. This background estimate

is further used for the background model development. This is done by dividing the image into blocks, and then motion between the corresponding blocks across the frames is calculated. If the block motion is under a threshold, then the block is declared as stationary. This process is repeated till all the blocks are declared as stationary. Median value is used for background updation . This method is preferred over mean, because mean value is highly susceptible to noise and local illumination variation. Pixel values are collected over time by a circular buffer. The current background estimate values are also added to the buffer. Median of the accumulated values forms the current background model. Foreground objects are extracted by using frame differencing method between the current frames and the current background model. For each pixel, two thresholds are defined. A minimum threshold for filtering the noisy pixels extracted due to small variation in intensity. A high threshold is used for identification of pixels having large intensity variation. Selective background update is implemented by the background pixels extracted by background subtraction method. Ghost suppression is also performed by this scheme. This occurs because selectivity is performed on detected moving objects rather than reasoning on single moving points. Shadow detection is done by using the HSV color space. It is assumed that shadows darken the underlying background but don't necessarily change its color. In other words the hue information is restored , whereas only the intensity is disturbed. The thresholds for this shadow suppression module are selected in such a way that the lower bound takes care of the darkening effect of the shadows on the background whereas the upper bound prevents the dark parts of the background to be falsely detect as shadows. Subsequently an object validation is used for removing the small moving objects due to motion in background. Joint information from color and gradient cues is used for performing this validation step. Finally a ghost suppression procedure is included.

Dual problem of saliency detection is equated with the background subtraction by Mahadevan et al. [39] The criterion for deciding whether a point belong to background is that actual background points are not salient by suitable comparison of object and background dynamics. Center-surround computation which measure

local feature contrast is used to define saliency locally. Saliency of location is defined using discriminant formulation. Discriminant power of a set of features with respect to the binary classification problem which opposes center to surround is used as saliency of a location. These features are modeled as dynamic textures which are spatiotemporal patches. This helps in making the scheme robust to highly dynamic backgrounds. It also accounts for motion and appearance. The main advantage of the resulting background subtraction algorithm is that it is fully unsupervised; therefore it doesn't require any training stage to learn the background parameters. It only depends on the relative disparity of motion between the center and the neighboring regions, which also makes the scheme resilient to camera motion.

Srivastava et al. used gaussianity test and shading model for proposing a novel scheme on background subtraction which emphasized on sudden illumination changes [40]. A hierarchical framework was used, which utilized block bases as well as pixel based processing. Intensity differences and intensity ratios of current frame and background model were used for block level classification. Foreground mask was extracted by pixel wise adaptive background differentiation on the foreground classified blocks. This method was resilient to sudden illumination changes and motion in cluttered background. It is assumed that the distribution of camera noise is spatially Gaussian. Gaussian test was implemented by using the intensity differences between the current frame and the background model. To impart the quality of robustness, shading model was imposed on the Gaussian test. This made the scheme resilient to sudden illumination change. Conventional background subtraction algorithms used intensity differencing for deciding whether a point belonged to the foreground or not. Here in this method, intensity ratios between the input frame and the already built background model is used for gaussianity test. In background model initialization, frame differencing is used to extract a background model. Selective update method is used for background model development. The main purpose of using a gaussianity test is to confirm whether the samples lie on a Gaussian curve or not. Camera noise which is assumed to be spatially Gaussian is also assumed to be temporally uncorrelated.

Therefore it is independent across the frames. So, after the frame differencing is done, the objects so detected will belong either to the Gaussian noise or actual objects. In other words, background pixels will have a distribution similar to Gaussian, while foreground pixels won't have a Gaussian distribution. This information helps in eliminating the Gaussian noise from the true foreground objects. This model works fine in normal scenarios, but fails miserably during scenarios of sudden illumination changes. This is because the assumption related to Gaussian distribution doesn't hold true when there are severe photometric distortions. When there is global illumination change, the increment in the intensities is not uniform throughout all the pixels. So the intensity ratios cannot be modeled as Gaussian. As a result, most of the pixels are falsely classified as object. To alleviate this problem, a shading model imposed on the gaussianity test is incorporated into the algorithm. However for this scheme to work, it is assumed that during the sudden illumination change across the frames, there is no object motion.

Jain et al. used a derivative model with a change model for developing a novel method for background subtraction which was illumination invariant [41]. Partial differentiation of the pixel location of a second order gray level surface model is done for determining the occurrence of a change. This is the fundamental idea behind the derivative model. Structural changes in the scene are detected by using a shading model. The prerequisite for application of the derivative model, is the presence of a surface model which is insensitive to illumination variations. Shading model efficiency directly depends on the shading coefficients. These coefficients are dependent on the physical surface of the object. They are not vulnerable to illumination changes. So, this property is exploited in making the algorithm robust to illumination changes. The main limitation of this approach is that it requires prior knowledge for calculation of shading coefficients. But in this algorithm absolute values of shading coefficient are not mandatory, only a detection of change in these coefficients is required. Ratio of intensities is exploited for this change detection. This model is also insensitive to noise.

Shoushtarian et al. used an invariant color filter and motion tracking technique

for object level classification [42]. The main advantage of this algorithm is that it is able to perform well in outdoor and indoor scenes. Moreover complex scenarios involving ghosts and objects becoming stationary are also handled properly. Background model updation is done selectively by using temporal averaging of frames. Temporal averaging of each specific color channel is done as long as occlusion by a moving object doesn't occur. These candidate background pixels are further used for the background model development. During occlusion by an object, the pixel intensity of the actual background should remain unchanged. So for the development of the background model, pixel values before occlusion and after occlusion are only considered. In this way, a proper estimate of the background is produced even if there is object motion in the input frames. Median value of the pixels is used for background updation. This is because it is more robust to noise and illumination variation as compared to mean. Pixel wise difference is taken between the current image and the background model for all the three channels. This difference is compared with their corresponding thresholds. If the condition doesn't satisfy for a single channel, then pixel is classified as foreground. Connected component analysis and morphological operation is applied subsequently to get actual objects ignoring the speckle noise. Normalized rgb color space is used for defining the color filter which has to be applied subsequently.

Apart from the challenge of sudden illumination change, the problem of shadows has been quite troublesome in developing an efficient background modelling scheme. Cucchiara et al. used HSV color space for shadow suppression [43]. Here it is assumed that shadows affect the brightness part of the image more as compared to the color information. There is similarity in texture between the shadow pixels and the corresponding background pixels. HSV color space is more resilient to illumination variation as compared to standard RGB color space. It is assumed that shadowed pixels have lower brightness but similar chromaticity. For simplicity, shadow detection and suppression is done after the possible moving objects have been extracted by the frame differencing technique. It is assumed that the intensity ratios between the shadow pixels and the corresponding background pixels would be nearly constant or rather

would not differ much. This ratio would be limited in a range of values. The difference between the hue components between the background model and the shadowed pixels would be limited by a threshold. Similar threshold would also be defined for saturation values. These conditions help in defining the shadow mask. Suitable thresholds take care of background noise being falsely detect as shadows as well as strong shadows which happen in scenarios of strong light. The actual parameters of the algorithm are selected empirically. Temporal median function is used for background model updation, so that it can adapt itself to the changing illumination conditions.

After surveying numerous papers based on background subtraction methods, interesting features can be seen common to all. Most of the algorithms have a background modeling phase, followed by object detection phase and finally filtering of irrelevant objects like cast shadows and background noise. The intricacy of the algorithm used incremented with the use of global threshold. Generally algorithms have to face challenges like photometric distortions, occlusion, complex object motion. The schemes which tried to solve these issues had high computational complexity. Therefore one has to trade off accuracy with computational cost. Here in this thesis, we proposed a scheme which has feasible computation cost with good performance.

Chapter 3

Proposed Work

3.1 Background

Our object detection algorithm is based on [11]. Here a local threshold based background subtraction method is used for object detection. There are two contributions done in this thesis. Firstly sudden illumination change scenarios were handled by foreground extraction using optical flow analysis. Secondly shadow suppression module was included for making the algorithm robust to shadows which were falsely classified as objects.

Conventional background subtraction algorithms failed to perform in scenarios of severe photometric distortions. Under the assumption that the object is in motion during the sudden illumination change and it has the largest motion in the scene, optical flow analysis is used to extract the foreground. Thresholding of the magnitude of motion vector is done for object extraction.

Typical background subtraction methods are plagued by the problem of shadows. Shadows are generally misclassified as objects. This leads to distortion of the object characteristics like area, centroid, shape and other features. In this thesis, shadow suppression module is incorporated into the algorithm. HSV color space model is used for this technique [43].

3.2 Proposed Detection Model

The proposed object detection model takes few initial frames to model the background. The foreground objects can be detected in any subsequent frame by comparing it with the developed model. The proposed model is capable enough to filter shadows associated with the object which may be falsely classified as foreground. The detected objects can be used to perform higher cognitive tasks like tracking or object analysis.

3.2.1 Background Modelling

In this phase, starting few frames are used for building the background model. For each pixel, a window is considered which stores the pixel values across certain number of frames. Classification of stationary or non-stationary pixels is done on the basis of their respective deviation from the mean of the pixel values gathered across the window. Subsequently all these stationary pixels are used for the development of the background model. Finally a range of values is defined for each background pixel.

3.2.2 Background Subtraction

The method used for extracting foreground objects is based on the local thresholding based background subtraction. Each background pixel has a range of values defined in the modeling phase. Among this range of values the maximum and minimum values are selected to form the thresholds for each pixel. Minimum and maximum thresholds are defined for each background pixel. These bounds alleviate the problem of shadows and over exposure. The minimum threshold takes care of faint shadows being falsely detected as objects. The maximum threshold suppresses glare so that it doesn't get misclassified as object. This also takes care of the periodic motion of background components.

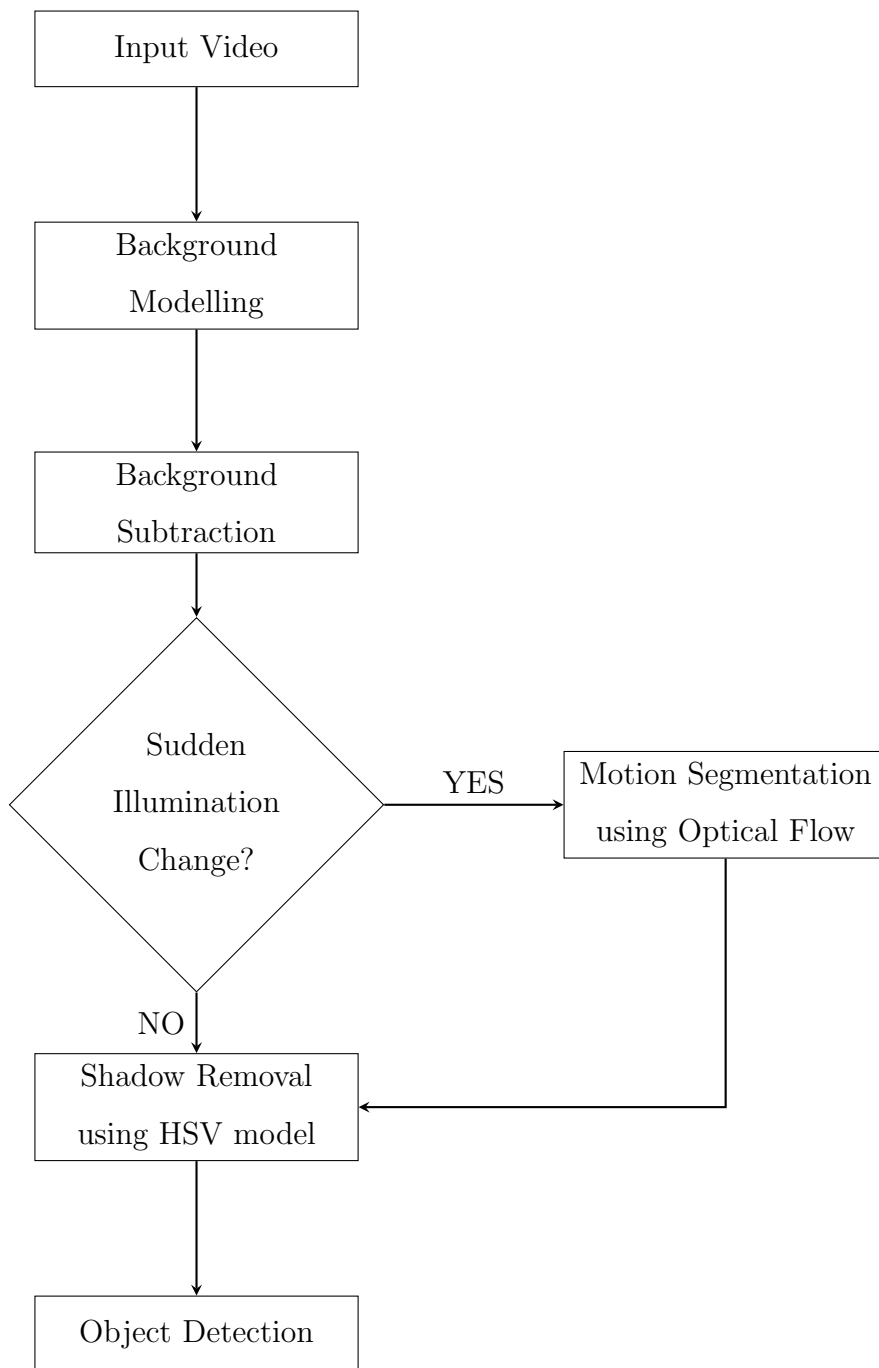


Figure 3.1: Proposed Detection Model

3.2.3 Motion Segmentation using Optical Flow

The previous scheme didn't give efficient performance during photometric distortions. In such case, it is quite challenging to develop an accurate background model which can extract the objects efficiently. In our thesis, optical flow is used to extract the moving regions of the scene during this sudden illumination change. Though optical flow works when the brightness constancy constraint is satisfied, using this method in such circumstances gives better result than the local thresholding based background subtraction. Calculation of optical flow is quite computational intensive. Optical flow calculation is only done during sudden photometric distortions, where majority of the pixels are falsely detected as foreground.

3.2.4 Shadow Removal using HSV color space

Though the local thresholding based background subtraction method suppresses the weak shadows in the background scene, it fails to do so in case of strong shadows. In such scenarios, HSV color space is utilized since it is more resilient to illumination variation as compared to RGB color space [43].

The details of the algorithm are described as below.

Algorithm 1: Development of Background model

```

1 Consider  $n$  initial frames as  $\{F_1, F_2 \dots F_n\}$ , where  $20 \leq n \leq 30$  begin
2   for  $k \leftarrow 1$  to  $n - (W - 1)$  do
3     for  $i \leftarrow 1$  to height of frame do
4       for  $j \leftarrow 1$  to width of frame do
5          $\vec{V} \leftarrow [f_k(i, j), f_{k+1}(i, j) \dots f_{k+(W-1)}(i, j)]$ 
6          $\sigma \leftarrow$  standard deviation of  $\vec{V}$ 
7          $D(p) \leftarrow |V(k + (\lfloor W \div 2 \rfloor)) - V(p)|$ ,
8         for each value of  $p = k + l$ ,
9         where  $l=0, \dots, (W-1)$  and  $l \neq \lfloor W \div 2 \rfloor$ 
10         $S \leftarrow$  sum of lowest  $\lfloor W \div 2 \rfloor$  values in  $\vec{D}$ 
11        if  $S \leq \lfloor W \div 2 \rfloor \times \sigma$  then
12          Label  $f_{k+\lfloor W \div 2 \rfloor}(i, j)$  as stationary
13        else
14          Label  $f_{k+\lfloor W \div 2 \rfloor}(i, j)$  as non-stationary
15      for  $i \leftarrow 1$  to height of frame do
16        for  $j \leftarrow 1$  to width of frame do
17           $M(i, j) = \min[f_s(i, j)]$  and,
18           $N(i, j) = \max[f_s(i, j)]$ ,
19          where  $s = \lfloor W \div 2 \rfloor \dots n - (\lfloor W \div 2 \rfloor)$  and  $f_s(i, j)$  is stationary
20 end

```

Algorithm 2: Background subtraction of a single frame

```

1 begin
2   for  $i \leftarrow 1$  to height of frame do
3     for  $j \leftarrow 1$  to width of frame do
4       Threshold  $T(i, j) = [M(i, j) + N(i, j)] \div C$ 
5        $T_L(i, j) = M(i, j) - T(i, j)$ 
6        $T_U(i, j) = M(i, j) + T(i, j)$ 
7       if  $T_L(i, j) \leq f(i, j) \leq T_U(i, j)$  then
8          $S_f(i, j) = 0$  //Background pixel
9       else
10         $S_f(i, j) = 1$  //Foreground pixel
11 end

```

Algorithm 3: Detection of sudden illumination change and subsequent object detection using optical flow analysis

```

1 begin
2   for  $i \leftarrow 1$  to height of frame do
3     for  $j \leftarrow 1$  to width of frame do
4       if  $S_f(i, j) = 1$  then
5         //If object pixel
6          $count = count + 1$  //Counting the total object pixels
7   Object Detection Percentage  $O_{dp} = count \div [(height) * (frame)]$ 
8   if  $O_{dp} \geq T_{dp}$  then
9     // if most of the pixels are detected as object
10     $S_f = \text{OpticalFlowThreshold}(S_f, S_{f-1})$ 
11 end

```

Algorithm 4: Shadow suppression module

```

1 begin
2   for  $i \leftarrow 1$  to height of frame do
3     for  $j \leftarrow 1$  to width of frame do
4       if  $S_f(i, j) = 1$  then
5         //If object pixel
6          $ratio = S_f^V(i, j) \div B_f^V(i, j)$ 
7          $diffS = S_f^S(i, j) - B_f^S(i, j)$ 
8          $diffH = |S_f^H(i, j) \div B_f^H(i, j)|$ 
9       if  $(\alpha \leq ratio \leq \beta) \wedge (diffS \leq T_S) \wedge (diffH \leq T_H)$  then
10        // shadow successfully detected
11         $S_f(i, j) = 0$  //Merging shadow pixel with background
12 end

```

Initially for the development of the background model, we need to find the stationary pixels in starting few frames. For each pixel, a window of frames (W) is considered. The pixel distribution is stored in a vector \vec{V} . If half of the pixels in that vector deviate from its mean within a threshold defined by half the window size times the standard deviation, then the center pixel is classified as stationary. In similar fashion, all the stationary pixels are accumulated. Subsequently for each stationary pixel, its maximum and minimum intensity range is evaluated. These values are used to define local thresholds for each pixel. The lower threshold takes care of faint shadows and low intensity background noise. While the upper

threshold takes care of the over exposure conditions.

Generally during sudden illumination change, this model fails to detect objects accurately. Most of the pixels are falsely detected as foreground. This condition is detected by evaluating the percentage of foreground extracted against a threshold T_{dp} which is reflected in the Algorithm 3. Upon detection, optical flow analysis is used to threshold the frames to extract the moving regions between them. This threshold information is again stored back to the frames S_f .

Shadow suppression is done by using the fact that HSV color space is more resilient to photometric distortions as compared to the standard RGB model. Shadow pixels have lower brightness but similar chromaticity as compared to the background pixels [43]. For all the foreground pixels detected, shadow detection method is applied. $S_f^V(i, j)$ represent the value component of the HSV pixel at the position (x, y) of the frame f of the video sequence S . Similarly hue and saturation components of the input frame and background frame are represented. Median value of the initial few frames is used as a proper estimate of the background model. Thresholds like α, β, T_S, T_H are computed empirically.

3.3 Results

Two challenging videos are used as inputs for this algorithm. They are :-

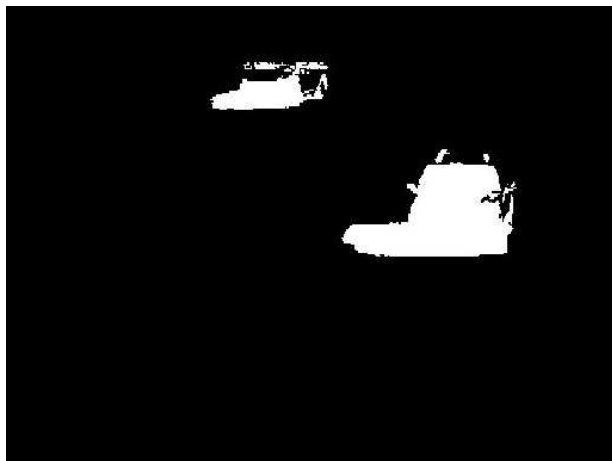
- **Traffic Sequence** This video was captured using a static camera in a traffic scene. The videos captured the motion of the cars along a straight road. Shadows are also present in the sequence, but they are not so prominent. The main property of this video is the presence of sudden illumination change. During normal scenarios the local threshold based background subtraction method gave proper result. But during photometric distortion, optical flow analysis was used for foreground detection. The figures convey the performance of the algorithm.
- **VipTraffic** This video was also captured using a static camera. The main property of the video is the presence of strong shadows which can be

easily be misclassified as objects. Our algorithm is able to suppress shadows efficiently. However few pixels of the objects are also misclassified as shadow pixels. The results are displayed as follows.

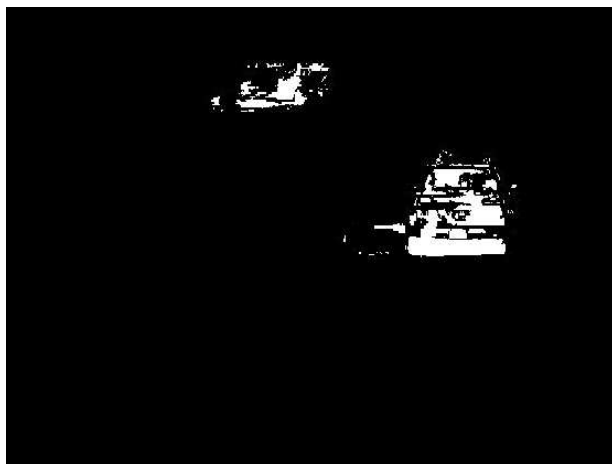
- **Light Switch** This challenging video has the scenes which portray severe photometric distortion [44]. The lighting of the scene changes drastically from light to dark within a span of two frames. Towards the end of the sequence, the illumination changes from dark to light. The sudden transition in room light is implemented using a light switch. There is a constant flicker in the screen of the monitor. This is accompanied by intermittent appearance of the user who operates the switch. Optical flow is also not able to extract the moving object effectively from this video sequence.



(a)



(b)

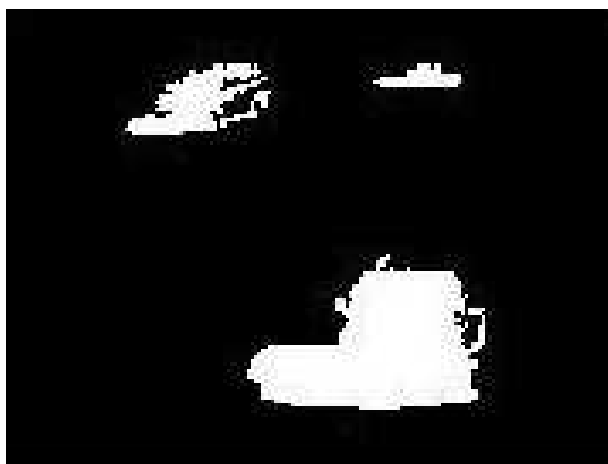


(c)

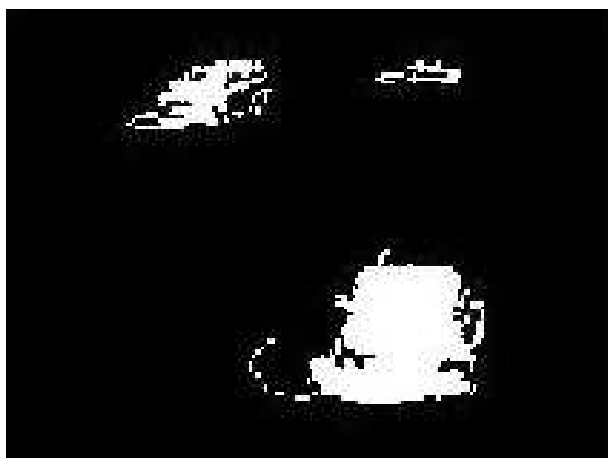
Figure 3.2: Shadow Suppression of *VipTraffic* image. (a) True Image, (b) Thresholded Image (c) Shadow suppressed



(a)



(b)

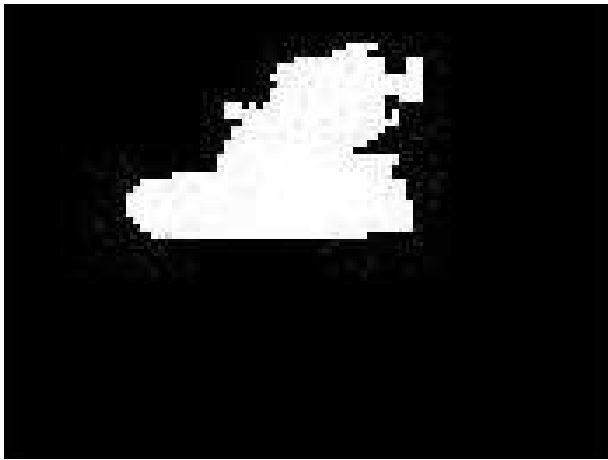


(c)

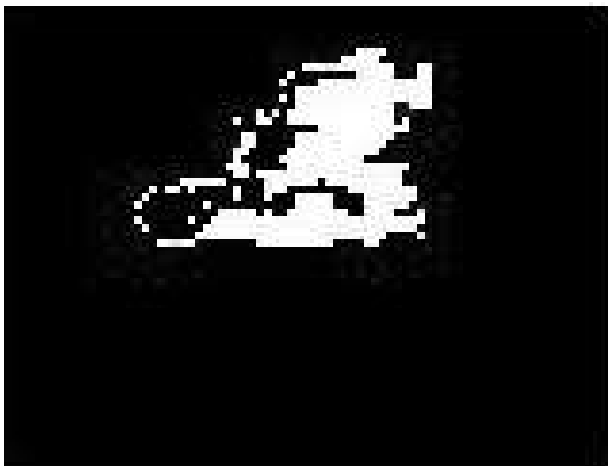
Figure 3.3: Shadow Suppression of *VipTraffic* image. (a) True Image, (b) Thresholded Image (c) Shadow suppressed



(a)



(b)



(c)

Figure 3.4: Shadow Suppression of *VipTraffic* image. (a) True Image, (b) Thresholded Image (c) Shadow suppressed

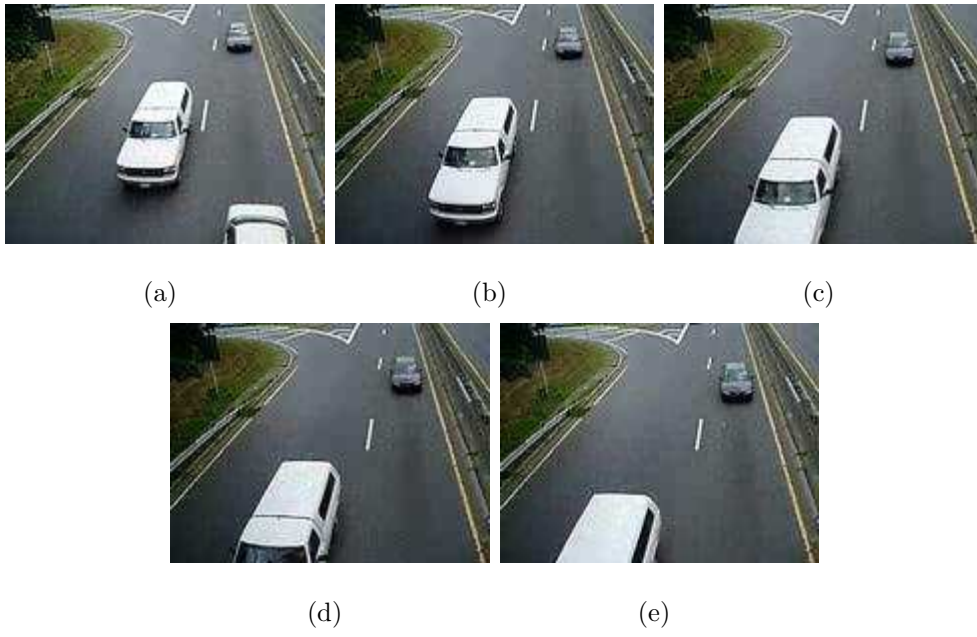


Figure 3.5: Input frames of *Traffic* sequence containing sudden illumination change

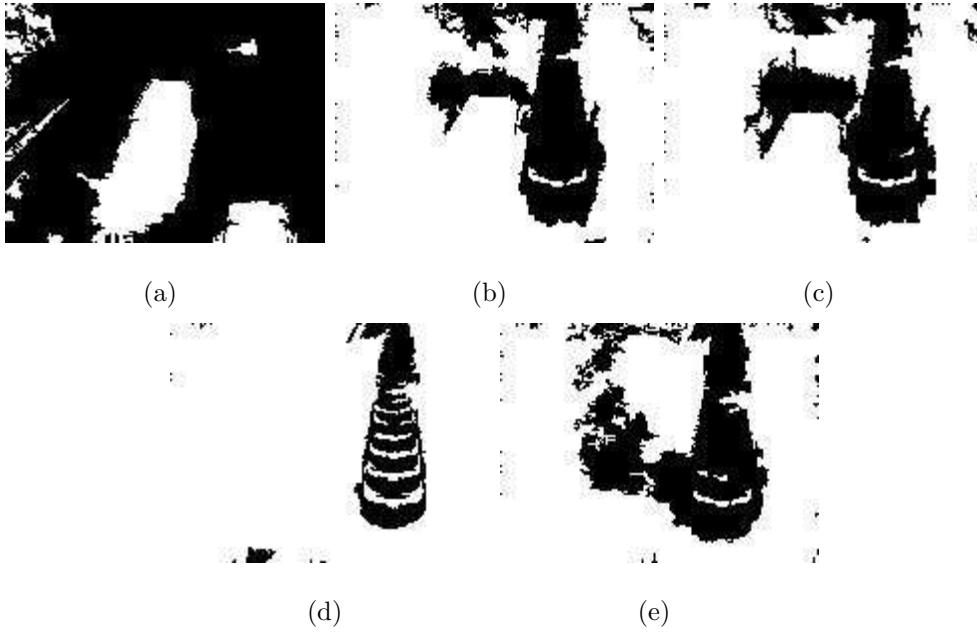


Figure 3.6: Inability of the existing model to extract objects accurately

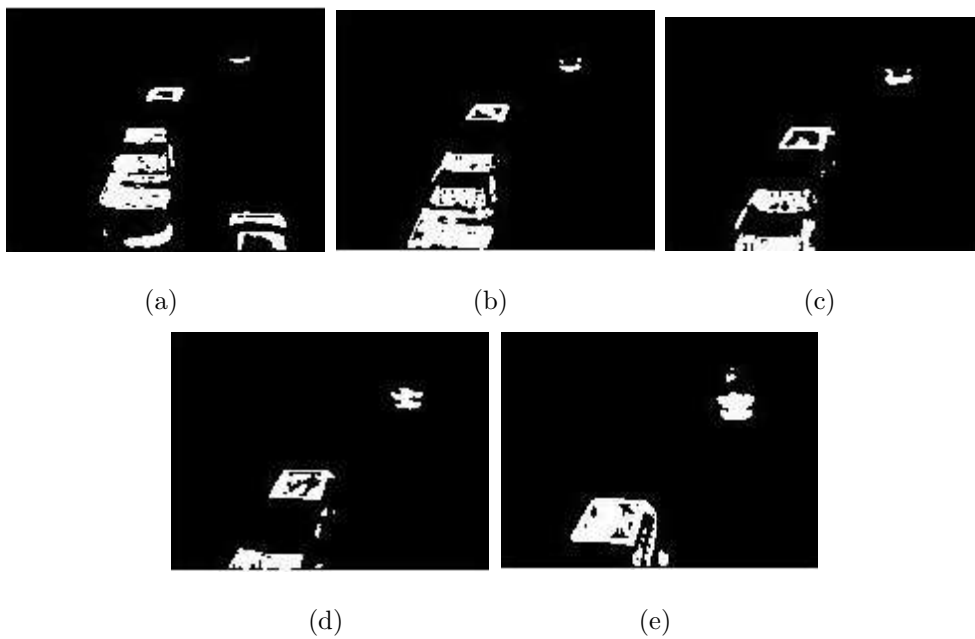


Figure 3.7: Object extraction using optical flow thresholding

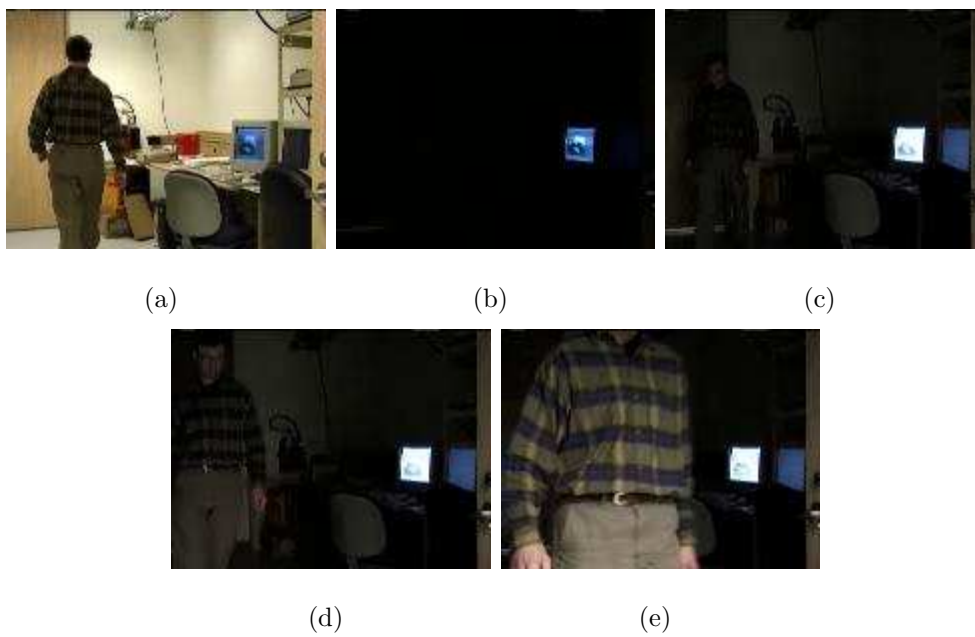


Figure 3.8: Original rgb frames of the Light Switch sequence

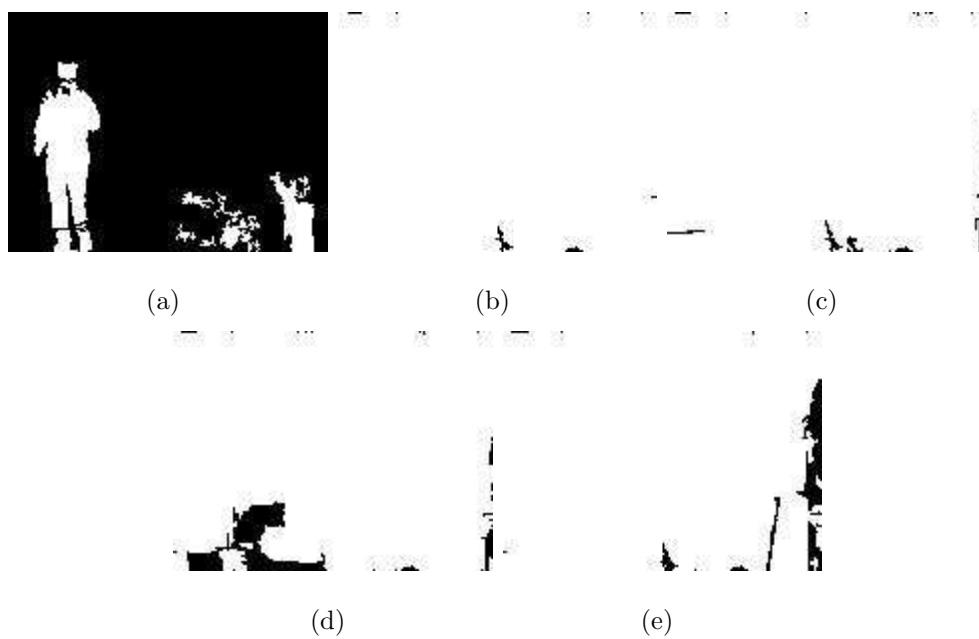


Figure 3.9: Inability of the existing model to detect object

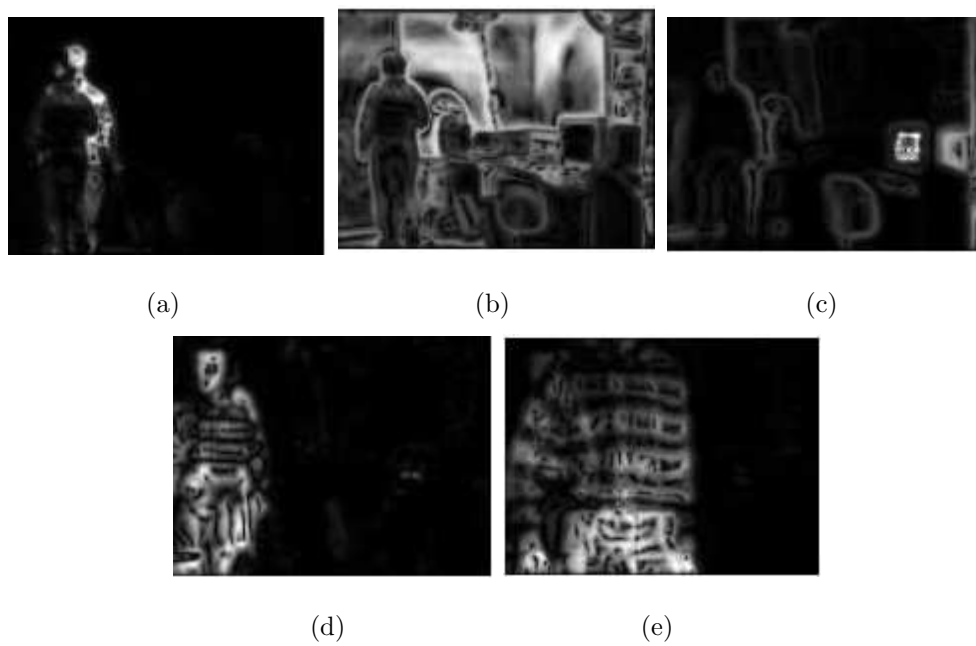


Figure 3.10: Result of optical flow analysis

Chapter 4

Conclusions and Future Scope

In the current generation of rapidly developing technologies, multimedia has deeply penetrated into all realms of life. One's daily routine has multiple encounters with multimedia services. One prominent reason for sudden upsurge of multimedia components is decrease in cost of technological gadgets like cameras and computers. Cameras have evolved rapidly throughout the century. The cost of image sensors has reduced drastically resulting in abundance of imaging devices. This resulted in huge accumulation of data in form of images and video clips. Extracting relevant information from this multimedia content was highly important. This need led to the development of object detection algorithms. Object detection methods play a vital role in the domain of surveillance. Accurate object detection and tracking led to higher cognitive tasks like event classification. Object tracking comprises two closely related processes; object detection followed by tracking of the detected objects. Object tracking is basically estimating the object's position using past information about its motion. Object tracking algorithms have gained quite popularity due to inexpensive yet high quality cameras and increasing need for automated video analysis. The surveillance system is the systematic method of behavior monitoring, actions or other changing information. Ideally the object detection system should have a static and uniform background from which the object may be extracted by using simple background subtraction operation. Object detection, object tracking and recognition are three

fundamental steps in any surveillance system. Video analysis comprises of relevant moving object detection, tracking of corresponding object from frame to frame and study of the motion history image to study its behavior.

In this thesis, a novel scheme for object detection in complex background scenes has been proposed. The input videos used have fixed backgrounds and static cameras. Initially median of few frames is evaluated for obtaining a proper estimate of the background. Local threshold based background subtraction is done for extracting objects from the video sequence. During sudden illumination changes, optical flow analysis is used for motion segmentation. It is assumed that during photometric distortions, the object is in motion. Subsequently shadow detection and suppression is done to the resulting thresholded image. Hue Saturation Value (HSV) color space model is used for shadow suppression. Visual measures convey the performance of the algorithm.

4.1 Scope for Further Research

Work done in this thesis can be extended into other broad domains. The efficiency of the algorithm proposed can be increased by using more robust features like intensity ratios constrained to a shading model. These features will give better result as compared to optical flow motion vectors. Moreover in scenarios of severe photometric distortions, our scheme is unable to detect objects effectively. In shadow suppression module, the parameters are computed empirically. This method could be modified to use adaptive values. Illumination invariant color space models may be used for better performance.

Bibliography

- [1] Rafael C Gonzalez, Richard Eugene Woods, and Steven L Eddins. *Digital image processing using MATLAB*. Pearson Education India, 2004.
- [2] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13, 2006.
- [3] Hans P Moravec. Visual mapping by a robot rover. In *Proceedings of the 6th international joint conference on Artificial intelligence-Volume 1*, pages 598–600. Morgan Kaufmann Publishers Inc., 1979.
- [4] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988.
- [5] Jianbo Shi and Carlo Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE, 1994.
- [6] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [7] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619, 2002.
- [8] Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm. In *ICML*, volume 96, pages 148–156, 1996.
- [9] Berthold K Horn and Brian G Schunck. Determining optical flow. In *1981 Technical Symposium East*, pages 319–331. International Society for Optics and Photonics, 1981.
- [10] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981.
- [11] Kalyan Kumar Hati, Pankaj Kumar Sa, and Banshidhar Majhi. Intensity range based background subtraction for effective object detection. *Signal Processing Letters, IEEE*, 20(8):759–762, 2013.

- [12] Dieter Koller, Joseph Weber, and Jitendra Malik. *Robust multiple car tracking with occlusion reasoning*. Springer, 1994.
- [13] Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. Wallflower: Principles and practice of background maintenance. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 255–261. IEEE, 1999.
- [14] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland. Pfinder: Real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):780–785, 1997.
- [15] Dieter Koller, J Weber, T Huang, J Malik, G Ogasawara, B Rao, and S Russell. Towards robust automatic traffic scene analysis in real-time. In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 126–131. IEEE, 1994.
- [16] BPL Lo and SA Velastin. Automatic congestion detection system for underground platforms. In *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on*, pages 158–161. IEEE, 2001.
- [17] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. Detecting moving objects, ghosts, and shadows in video streams. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(10):1337–1342, 2003.
- [18] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE, 1999.
- [19] Ahmed Elgammal, David Harwood, and Larry Davis. Non-parametric model for background subtraction. In *Computer Vision ECCV 2000*, pages 751–767. Springer, 2000.
- [20] Massimo Piccardi. Background subtraction techniques: a review. In *Systems, man and cybernetics, 2004 IEEE international conference on*, volume 4, pages 3099–3104. IEEE, 2004.
- [21] Makito Seki, Toshikazu Wada, Hideto Fujiwara, and Kazuhiko Sumi. Background subtraction based on cooccurrence of image variations. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–65. IEEE, 2003.
- [22] P Wayne Power and Johann A Schoonees. Understanding background mixture models for foreground segmentation. In *Proceedings image and vision computing New Zealand*, volume 2002, pages 10–11, 2002.

-
- [23] Nuria M Oliver, Barbara Rosario, and Alex P Pentland. A bayesian computer vision system for modeling human interactions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):831–843, 2000.
 - [24] Du-Ming Tsai and Shia-Chih Lai. Independent component analysis-based background subtraction for indoor surveillance. *Image Processing, IEEE Transactions on*, 18(1):158–167, 2009.
 - [25] Horng-Horng Lin, Tyng-Luh Liu, and Jen-Hui Chuang. Learning a scene background model via classification. *Signal Processing, IEEE Transactions on*, 57(5):1641–1654, 2009.
 - [26] Lucia Maddalena and Alfredo Petrosino. A self-organizing approach to background subtraction for visual surveillance applications. *Image Processing, IEEE Transactions on*, 17(7):1168–1177, 2008.
 - [27] Ismail Haritaoglu, David Harwood, and Larry S. Davis. W 4: Real-time surveillance of people and their activities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):809–830, 2000.
 - [28] Daniel Gutchess, M Trajkovics, Eric Cohen-Solal, Damian Lyons, and Anil K Jain. A background model initialization algorithm for video surveillance. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 733–740. IEEE, 2001.
 - [29] J Cezar Silveira Jacques, Claudio Rosito Jung, and S Raupp Musse. Background subtraction and shadow detection in grayscale video sequences. In *Computer Graphics and Image Processing, 2005. SIBGRAPI 2005. 18th Brazilian Symposium on*, pages 189–196. IEEE, 2005.
 - [30] Cláudio Rosito Jung. Efficient background subtraction and shadow removal for monochromatic video sequences. *IEEE Transactions on Multimedia*, 11(3):571–577, 2009.
 - [31] Olivier Barnich and Marc Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *Image Processing, IEEE Transactions on*, 20(6):1709–1724, 2011.
 - [32] Wonjun Kim and Changick Kim. Background subtraction for dynamic texture scenes using fuzzy color histograms. *Signal Processing Letters, IEEE*, 19(3):127–130, 2012.
 - [33] Vikas Reddy, Conrad Sanderson, and Brian C Lovell. Improved foreground detection via block-based classifier cascade with probabilistic decision integration. 2013.
 - [34] Julien Pilet, Christoph Strecha, and Pascal Fua. Making background subtraction robust to sudden illumination changes. In *Computer Vision–ECCV 2008*, pages 567–580. Springer, 2008.

-
- [35] JinMin Choi, Hyung Jin Chang, Yung Jun Yoo, and Jin Young Choi. Robust moving object detection against fast illumination change. *Computer Vision and Image Understanding*, 116(2):179–193, 2012.
- [36] Kyungnam Kim, Thanarat H Chalidabhongse, David Harwood, and Larry Davis. Real-time foreground–background segmentation using codebook model. *Real-time imaging*, 11(3):172–185, 2005.
- [37] Luc Vosters, Caifeng Shan, and Tommaso Gritti. Real-time robust background subtraction under rapidly changing illumination conditions. *Image and Vision Computing*, 30(12):1004–1015, 2012.
- [38] Simone Calderara, Rudy Melli, Andrea Prati, and Rita Cucchiara. Reliable background suppression for complex scenes. In *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, pages 211–214. ACM, 2006.
- [39] Vijay Mahadevan and Nuno Vasconcelos. Background subtraction in highly dynamic scenes. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–6. IEEE, 2008.
- [40] Ka Ki Ng, Satyam Srivastava, and Edward J Delp. Foreground segmentation with sudden illumination changes using a shading model and a gaussianity test. In *Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on*, pages 236–240. IEEE, 2011.
- [41] Kurt Skifstad and Ramesh Jain. Illumination independent change detection for real world image sequences. *Computer Vision, Graphics, and Image Processing*, 46(3):387–399, 1989.
- [42] Bijan Shoushtarian and Helmut E Bez. A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking. *Pattern Recognition Letters*, 26(1):5–26, 2005.
- [43] Rita Cucchiara, Costantino Grana, Massimo Piccardi, Andrea Prati, and Stefano Sirotti. Improving shadow suppression in moving object detection with hsv color information. In *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, pages 334–339. IEEE, 2001.
- [44] Wallflower. Test images for wallflower paper. <http://research.microsoft.com/en-us/um/people/jckrumm/WallFlower/TestImages.htm>.